

生理・代謝機能ポテンシャル評価システム Genomaple の ES4 への移植

課題責任者

古市 幹人 海洋研究開発機構 付加価値情報創生部門 数理科学・先端技術研究開発センター

著者

山岸 保子*¹, 古市 幹人*¹, 石渡 千恵子*¹, 高見 英人*^{1,2}

*¹ 海洋研究開発機構 付加価値情報創生部門 数理科学・先端技術研究開発センター, *² 株式会社 C02 資源化研究所

キーワード：微生物, メタゲノム, 生理代謝機能, 機能ポテンシャル解析, web application

1. 序論

Genomaple は海洋研究開発機構が開発した、ゲノムデータベースを検索するツールであり、データベースを検索することで、モジュールに基づく生理・代謝機能の有無を判定する。すなわち、ゲノム・メタゲノム配列から、様々な環境に生息する微生物（群）が有する生理・代謝機能ポテンシャルを評価する解析システムである。本システムは、2018年9月から、機構の所内向け大型計算機 DA システムに搭載したものを、webUI を提供することで外部に一般公開していた。しかし、2021年に機構で発生したセキュリティインシデントによって公開中断という事態に陥った。公開した Genomaple は、ブラウザからインターネットを通じて使用可能という利便性により、公開から中断まで2年半という短い期間にもかかわらず、46カ国、約400名超の研究者により月平均200件の利用があった。また、Genomaple システムの開発において発表された解析原理に関する論文[1]、並びに解析システムの自動化や解析例に関する論文[2]、さらにシステムの高速化に関する論文[3]は、現在までに数多くの研究論文で引用されており、ゲノム・メタゲノム研究者にとって、Genomaple が有用なツールであると世界的に認知されていることがわかる。よって本課題では、Genomaple 再公開と、さらに、地球シミュレータ (ES4) のウェブアプリケーションのエンジン化という新たな利用法の提示を目標とした。そのため、ES4 に Genomaple を移植して、Web 関連システムを含めて様々な改修を行った。その結果、2023年7月に Genomaple-ES4 として再び公開することが出来た。図1は、Genomaple-ES4 の、ゲノム・メタゲノム解析研究における役割の概念図である。その後、Genomaple は開発モードから運用モードに切り替わったが、残念ながらデータベースの使用に関して問題が発生し、解決に至ることが出来なかった。そして、改修・改良を加えて再公開した Genomaple-ES4 も、2024年10月にサービスの提供を終了した。

2. Genomaple の ES4 への移植

本課題では Genomaple を再公開するために、ES4 へ移植を実施したが、以下の二点を重要視した。

1. ES4 の GPU 活用
2. セキュリティ強化

ES4 はマルチアーキテクチャが特性であり、特に GPU の活用が望まれる。そのため、CPU ベースが基本であったシステムに GPU を効率的に利用するための機能を追加開発した。その結果、データ処理速度は、DA システムと比較して約2倍の高速化を達成し、また同時処理が可能なデータ数の限界値も300万配列から1000万配列と大幅に増加させることが出来た。Genomaple-ES4 は、外部公開を前提としているため、セキュリティの強化は最優先事項に当たる。そのため、まずユーザーのログインに対して二要素認証システムを導入することで不正ログイン対策を行なった。またシステム構成として中継サーバを導入して、Web システムを搭載したフロントサーバと ES4 の間に設置することで、外部公開用領域である DMZ と ES4 のネットワーク領域を完全に遮断した。また各サーバの運用に関しても厳しい規制を行なった。重要視した二点以外にも様々な改修を行い、特にユーザーの利便性を高める機能の追加開発を実施した。例えば、システムに投入された各クエリ配列がどのような代謝・生物機能に関与するのかを、よりわかりやすく整理するとともに、ユーザがそれらの関連データファイルをダウンロード出来るようにした。UI についても改修を行い、計算結果の表示も改良した。



図1. Genomaple-ES4 のゲノム・メタゲノム解析における役割

3. 再公開後の利用実績

前述の通り、Genomaple を ES4 に移植して、新たに Genomaple-ES4 として2023年7月に外部に一般公開した。

再公開月の新規ユーザー数は約10人、その後も大体一月あたり2~5名の新規ユーザー登録があり、順調にユーザー数は増加した。しかし、データベースに関する問題が発覚して、2024年3月に運用を一時中断した。再公開から中断までの新規登録ユーザーは約40人、総ユーザー数は約300人弱であった。ES4の定期メンテナンスなどが無かった月の利用実績は約100~140件であった。DAシステム搭載時は約400名超ユーザーにより月平均200件の利用があったが、セキュリティインシデントによる中断期間中に非アクティブとなったユーザーを整理した結果、ユーザーの総数が100名ほど減少したこと考慮すると、Genomagle-ES4の利用実績はDA搭載時と遜色なく、サービス利用に関しての需要は変わらなかったと考えられる。2024年3月に発生したデータベースの問題は解決が困難であることがわかり、サービスの終了が決定された。既存ユーザーに対しては、後処理や投稿論文対策のため、2024年7月から10月まで期間限定で再公開した。この間にも、新規利用希望者からの問い合わせがあり、また毎月約30から80件の利用があった。サービスは終了したものの、ゲノム・メタゲノム研究者にとって需要のあるシステムであったと考えられる。

4. ES4の新たな活用法の提示

残念ながらサービスは終了したGenomagle-ES4であるが、ES4をWebアプリケーションのバックエンドサーバとして利用した初のシステムであった。すなわちES4の新しい利用方法を提示したシステムと言える。一方で、ES4をバックエンドサーバとして利用したことにより、システム全体の構成は複雑化し、複数のネットワーク領域を横断することになった。そのため、Genomagle-ES4は、挙動が度々不安定になることがあり、その原因は残念ながら短い運用期間中に明らかにすることは出来なかった。しかし、開発時、および運用時に蓄積した知見は、今後同様にES4をWebアプリケーションのバックエンドとして利用するシステムの開発に活かすことが可能である。

5. Genomagleによるバイオサイエンスの発展への貢献

DAシステム上での稼働を含めると約4年弱公開されたGenomagleシステムを利用して、多くの研究成果が発表された(例えば[4-5])。またGenomagle-ES4本体に関しても2024年2月に出版された書籍[6]に記事が掲載された。運用を停止したGenomagleではあるが、運用中の利用実績から、その需要は失われておらず、自然環境や腸内環境などのメタゲノム配列とメタゲノム配列から再構築された個別生物種のゲノム配列解析や比較解析には必要なシステムであると考えられる。Genomagleは100万リードを超えるショットガンメタゲノム配列、個別ゲノム、MAG (metagenome assembled genomes) の機能解析を可能にするシステムであったが、後者の個別ゲノムとMAGの機能解析について、Genomagleを利用した解析システムが国立研究開発法人基礎生物学研究所により開発された[7]。

このことから、本システムのバイオサイエンスにおける重要性が示唆される。短い期間ではあったが、Genomagle-ES4によるゲノム・メタゲノム解析は、バイオサイエンスの発展に貢献したと考えられる。

謝辞

GenomagleシステムのES4への移植並びにWeb関連システムの再構築、そして外部公開と運用は、当機構の付加価値情報創生部門・地球情報科学技術センターの中川剛史氏と石黒駿氏、および情報セキュリティ・システム部情報セキュリティ統括課の古川向直氏並びに豊村鉄男氏の多大な協力を得て実施した。ここに記し深く感謝する。

文献

- [1] H. Takami et al., "Evaluation method for potential functionome harbored in genome and metagenome", *BMC Genomics* 13, 699 (2012)
- [2] H. Takami et al., "An automated system for evaluation of the potential functionome: MAPLE version 2.1.0", *DNA Res.* 23, 467-475 (2016).
- [3] W. Arai et al., "MAPLE 2.3.0: An improved system for evaluating the functionomes of genomes and metagenomes", *Biosci. Biotech. Biochem.* 82(9), 1515-1517 (2018).
- [4] T. Shiozaki et al., "Distribution and survival strategies of endemic and cosmopolitan diazotrophs in the Arctic Ocean", *ISME J* 17, 1340-1350 (2023).
- [5] S. Jiang et al., "Variations in Physiology and Genomic Function of *Prochlorococcus* Across the Eastern Indian Ocean", *J. Geophys. Res. Oceans* 128(10), 2169-9275 (2023).
- [6] H. Takami, "Functional Microbial Diversity: Functional Genomics and Metagenomics Using Genomagle", *Microbial Diversity in the Genomic Era: Functional Diversity and Community Analysis*, 2nd ed., S. Das and H. R. Dash, ed., Academic Press, Cambridge, Massachusetts, 2024.
- [7] I. Uchiyama et al., "MBGD: Microbial Genome Database for Comparative Analysis Featuring Enhanced Functionality to Characterize Gene and Genome Functions Through Large-scale Orthology Analysis", *J. of Molecular Biology*, <https://doi.org/10.1016/j.jmb.2025.168957>, 2025.